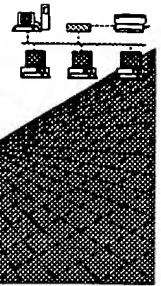


Roner

BIOTECHNOLOGY
SYSTEMS
BRANCH



RAW SEQUENCE LISTING **ERROR REPORT**

The Biotechnology Systems Branch of the Scientific and Technical Information Center (STIC) detected errors when processing the following CRF diskette:

Application Serial Number: 09/051,670
Art Unit / Team No. : 1646
Date Processed by STIC: 12/21/98

THE ATTACHED PRINTOUT EXPLAINS THE ERRORS DETECTED.

PLEASE BE SURE TO FORWARD THIS INFORMATION TO THE APPLICANTS BY EITHER:

1) INCLUDING A COPY OF THIS PRINTOUT IN YOUR NEXT COMMUNICATION TO THE APPLICANTS ALONG WITH A NOTICE TO COMPLY or,

2) CALLING APPLICANTS AND FAXING THEM A COPY OF THE PRINTOUT WITH A NOTICE TO COMPLY

THIS WILL INSURE THAT THE NEXT SUBMISSION RECEIVED FROM THEM WILL BE ERROR FREE.

IF YOU HAVE ANY FURTHER QUESTIONS, PLEASE CALL:

ARTI SHAH 703-308-4212

Raw Sequence Listing Error Summary

ERROR DETECTED SUGGESTED CORRECTION

SERIAL NUMBER: 09/051,670

ATTN: NEW RULES CASES: PLEASE DISREGARD ENGLISH "ALPHA" HEADERS, WHICH WERE INSERTED BY PTO SOFTWARE

- 1 Wrapped Nucleics The number/text at the end of each line "wrapped" down to the next line.
This may occur if your file was retrieved in a word processor after creating it.
Please adjust your right margin to .3, as this will prevent "wrapping".
- 2 Wrapped Aminos The amino acid number/text at the end of each line "wrapped" down to the next line.
This may occur if your file was retrieved in a word processor after creating it.
Please adjust your right margin to .3, as this will prevent "wrapping".
- 3 Incorrect Line Length The rules require that a line not exceed 72 characters in length. This includes spaces.
All text must be visible on page.
- 4 Misaligned Amino Acid The numbering under each 5th amino acid is misaligned. This may be caused by the use of tabs
Numbering between the numbering. It is recommended to delete any tabs and uses spacing between the numbers.
- 5 Non-ASCII This file was not saved in ASCII (DOS) text, as required by the Sequence Rules.
Please ensure your subsequent submission is saved in ASCII text so that it can be processed.
- 6 Variable Length Sequence(s) contain n's or Xaa's which represented more than one residue.
As per the rules, each n or Xaa can only represent a single residue.
Please present the maximum number of each residue having variable length and
indicate in the (ix) features section that some may be missing.
- 7 Wrong Designation Sequence(s) contain amino acid or nucleic acid designators which are not standard
representations as per the Sequence Rules (Please refer to paragraph 1.822)
- 8 Skipped Sequences Sequence(s) missing. If intentional, please use the following format for each skipped sequence:
(OLD RULES) (2) INFORMATION FOR SEQ ID NO:X:
 (1) SEQUENCE CHARACTERISTICS:(Do not insert any headings under "SEQUENCE CHARACTERISTICS")
 (x1) SEQUENCE DESCRIPTION:SEQ ID NO:X:
 This sequence is intentionally skipped

Please also adjust the "(iii) NUMBER OF SEQUENCES:" response to include the skipped sequence(s).
- 9 Skipped Sequences Sequence(s) missing. If intentional, please use the following format for each skipped sequence.
(NEW RULES) <210> sequence id number
 <400> sequence id number
 000
- 10 Use of N's or Xaa's Use of N's and/or Xaa's have been detected in the Sequence Listing.
(NEW RULES) Use of <220> to <223> is MANDATORY if n's or Xaa's are present.
- 11 Use of <213>Organism Sequence(s) are missing this mandatory field or its response.
(NEW RULES)
- 12 Use of <220>Feature Sequence(s) are missing the <220>Feature and associated headings.
(NEW RULES) Use of <220> to <223> is MANDATORY if <213>ORGANISM is "Artificial" or "Unknown"
 (See "Federal Register," 6/01/98, Vol. 63, No. 104, pp. 29631-32)
 (Sec. 1.823 of new Sequence Rules)
- 13 Wrong F mat File submitted was in the alphabetical heading format of the Old Sequence Rules. This is invalid since the
 "Requirements for Patent Applications Containing Nucleotide Sequence and/or Amino Acid Disclosures"
 Federal Register Notice, Vol. 63, No. 104, June 1, 1998, p. 29620
 applies to applications filed on or after July 1, 1998.
 AKS-Biotechnology Systems Branch- 7/10/98

(1) GENERAL INFORMATION:

- (i) Applicant: SNOW BRAND MILK PRODUCTS CO.,
 (ii) Title of the Invention: Novel DNA and process for preparing protein using the (DN)
 (iii) Number of Sequences: 4

(2) INFORMATION FOR SEQ ID NO: 1:

- (A) Length of sequence: 1316
 (B) Sequence type: nucleic acid
 (C) Strandedness: double
 (D) Topology: linear
 (i) Molecule type: genomic DNA

(v) CORRESPONDENCE ADDRESS:

- (A) ADDRESSEE:
 (B) STREET:
 (C) CITY:
 (D) STATE:
 (E) COUNTRY:
 (F) ZIP:

(vi) COMPUTER READABLE FORM:

- (A) MEDIUM TYPE:
 (B) COMPUTER:
 (C) OPERATING SYSTEM:
 (D) SOFTWARE:

(vii) CURRENT APPLICATION DATA:

- (A) APPLICATION NUMBER: ic DNA-1)
 (B) FILING DATE:
 (C) CLASSIFICATION:

These MANDATORY headings and responses are needed. All text must appear on page 3 (see item 3 on Envr)

Does Not Comply Summary Sheet
 Corrected Diskette Needed

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:

CTGGAGACAT ATAATTGAA CACTTGGGCC TGATGGGGAA GCAGCTCTGC AGGGACTTTT 60
 TCAGCCATCT GTAAACAATT TCAGTGGCAA CCCGCGAACT GTAATCCATG AATGGGACCA 120
 CACTTTACAA GTCATCAAGT CTAATTCTA GACCAGGGAA TTAATGGGGG AGACAGCGAA 180
 CCCTAGAGCA AAGTGCCAAA CTTCTGTCTA TAGCTTGAGG CTAGTGGAAA GACCTCGAGG 240
 AGGCTACTCC AGAAGTTCAG CGCGTAGGAA GCTCCGATAC CAATAGCCCT TTGATGATGG 300
 TGGGGTTGGT GAAGGGAACA GTGCTCCGCA AGGTTATCCC TGCCCCAGGC AGTCCAATTT 360
 TCACTCTGCA GATTCTCTCT GGCTCTAACT ACCCCAGATA ACAAGGAGTG AATGCAGAAT 420
 AGCAGGGGCT TTAGGGCCAA TCAGACATTA GTTAGAAAAA TTCCTACTAC ATGGTTTATG 480
 TAAACTTGAA GATGAATGAT TGCGAATCC CCGAAAAGGG CTCAGACAAAT GCCATGCATA 540
 AAGAGGGGCC CTGTAATTTG AGGTTTCAGA ACCCGAAGTG AAGGGGTCAG GCAGCCGGGT 600
 ACGGCGGAAA CTCACAGCTT TCGCCAGCGC AGAGGACAAA GGTCTGGGAC ACACTCCAAC 660
 TGCGTCCGGA TCTTGGCTGG ATCGGACTCT CAGGGTGGAG GAGACACAAG CACAGCAGCT 720
 GCCAGCGGTG TGCCAGCCCT TCCCACCGCT GGTCCCAGCT GCCAGGAGGC TGGCCGCTGG 780
 CGGGAAGGGG CCGGGAAGCC TCAGAGCCCC GCGGAGACAG CAGCCGCCTT GTTCCTCAGC 840
 CCGGTGGCTT TTTTTCCTCC TGCTCTCCCA GGGGACAGAC ACCACCGCCC CACCCCTCAC 900
 GCCCCACCTC CCTGGGGGAT CCTTTCGCC CCAGCCCTGA AAGCGTTAAT CCTGGAGCTT 960
 TCTGCACACC CCCCAGCCGC TCCCGCCCAA GCTTCTTAAA AAAGAAAGGT GCAAAGTTTG 1020
 GTCCAGGATA GAAAAATGAC TGATCAAAGG CAGGCGATAC TTCCTGTTGC CGGGACGCTA 1080
 TATATAACGT GATGAGCGCA CGGGCTGCGG AGACGCACCG GAGCGCTCGC CCAGCCGCCG 1140
 CCTCCAAGCC CCTGAGCTTT CCGGGGACCA CA ATG AAC AAG TTG CTG TGC TGC @ 1193

Met Asn Lys Leu Leu Cys Cys
 20
 GCG CTC GTG GTAAGTCCCT GGGCCAGCCG ACGGGTGCCC GGCGCCTGGG @ 1242
 Ala Leu Val

GAGGCTGCTG CCACCTGGTC TCCCAACCTC CCAGCGGACC GGCGGGGAGA AGGCTCCACT 1302
 CGCTCCCTCC CAGC @ 1316

(2) INFORMATION FOR SEQ ID NO: 2:

- (A) Length of sequence: 9898
 (B) Sequence type: nucleic acid
 (C) Strandedness: double
 (D) Topology: linear
 (i) Molecule type: genomic DNA (human OCIF genomic DNA-2)

all headings MUST be in upper-case letters

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

GCTTACTTTG TGCCAAATCT CATTAGGCTT AAGGTAATAC AGGACTTTGA GTCAAATGAT 60
 ACTGTTGCAC ATAAGAACAA ACCTATTTTC ATGCTAAGAT GATGCCACTG TGTTCTTTTC 120
 TCCTTCTAG TTT CTG GAC ATC TCC ATT AAG TGG ACC ACC CAG GAA ACG TTT 171
 Phe Leu Asp Ile Ser Ile Lys Trp Thr Thr Gln Glu Thr Phe
 10 50 100

CCT CCA AAG TAC CTT CAT TAT GAC GAA GAA ACC TCT CAT CAG CTG TTG @ 219
 Pro Pro Lys Tyr Leu His Tyr Asp Glu Glu Thr Ser His Gln Leu Leu

TGT GAC AAA TGT CCT CCT GGT ACC TAC CTA AAA CAA CAC TGT ACA GCA 267
 Cys Asp Lys Cys Pro Pro Gly Thr Tyr Leu Lys Gln His Cys Thr Ala 35
 20 25 30

AAG TGG AAG ACC GTG TGC GCC CCT TGC CCT GAC CAC TAC TAC ACA GAC 315
 Lys Trp Lys Thr Val Cys Ala Pro Cys Pro Asp His Tyr Tyr Thr Asp 50
 40 45 50

AGC TGG CAC ACC AGT GAC GAG TGT CTA TAC TGC AGC CCC GTG TGC AAG 363
 Ser Trp His Thr Ser Asp Glu Cys Leu Tyr Cys Ser Pro Val Cys Lys 65
 55 60 65

GAG CTG CAG TAC GTC AAG CAG GAG TGC AAT CGC ACC CAC AAC CGC GTG 411
 Glu Leu Gln Tyr Val Lys Gln Glu Cys Asn Arg Thr His Asn Arg Val 80
 70 75 80

TGC GAA TGC AAG GAA GGG CGC TAC CTT GAG ATA GAG TTC TGC TTG AAA 459
 Cys Glu Cys Lys Glu Gly Arg Tyr Leu Glu Ile Glu Phe Cys Leu Lys 95
 85 90 95

CAT AGG AGC TGC CCT CCT GGA TTT GGA GTG GTG CAA GCT G GTACGTGTCA 509
 His Arg Ser Cys Pro Pro Gly Phe Gly Val Val Gln Ala 110
 100 105 110

ATGTGCAGCA AAATTAATTA GGATCATGCA AAGTCAGATA GTTGTGACAG TTTAGGAGAA 569
 CACTTTTGTGTT CTGATGACAT TATAGGATAG CAAATTGCAA AGGTAATGAA ACCTGCCAGG 629
 TAGGTACTAT GTGTCTGGAG TGCTTCCAAA GGACCATTGC TCAGAGGAAT ACTTTGCCAC 689
 TACAGGGCAA TTTAATGACA AATCTCAAAT GCAGCAAATT ATTCTCTCAT GAGATGCATG 749
 ATGGTTTTTT TTTTTTTTTT TAAAGAAACA AACTCAAGTT GCACTATTGA TAGTTGATCT 809
 ATACCTCTAT ATTTCACTTC AGCATGGACA CCTTCAAACCT GCAGCACTTT TTGACAAACA 869
 TCAGAAATGT TAATTTATAC CAAGAGAGTA ATTATGCTCA TATTAATGAG ACTCTGGAGT 929
 GCTAACAATA AGCAGTTATA ATTAATTATG TAAAAAATGA GAATGGTGAG GGGAAATGCA 989
 TTTTATTATT AAAAACAAGG CTAGTTCTTC CTTTAGCATG GGAGCTGAGT GTTTGGGAGG 1049
 GTAAGGACTA TAGCAGAATC TCTTCAATGA GCTTATTCTT TATCTTAGAC AAAACAGATT 1109
 GTCAAGCCAA GAGCAAGCAC TTGCCTATAA ACCAAGTGCT TTCTCTTTTG CATTTTGAAC 1169
 AGCATTTGGTC AGGGCTCATG TGTATTGAAT CTTTTAAACC AGTAACCCAC GTTTTTTTTC 1229
 TGCCACATTT GCGAAGCTTC AGTGCAGCCT ATAACCTTTC ATAGCTTGAG AAAATTAAGA 1289
 GTATCCACTT ACTTAGATGG AAGAAGTAAT CAGTATAGAT TCTGATGACT CAGTTTGAAG 1349
 CAGTGTCTTCT CAACTGAAGC CCTGCTGATA TTTTAAGAAA TATCTGGATT CCTAGGCTGG 1409
 ACTCCTTTTT GTGGGCAGCT GTCCTGCGCA TTGTAGAATT TTGGCAGCAC CCCTGGACTC 1469
 TAGCCACTAG ATACCAATAG CAGTCCTTCC CCCATGTGAC AGCCAAAAAT GTCTTCAGAC 1529
 ACTGTCAAAT GTCGCCAGGT GGCAAAATCA CTCCTGCTTG AGAACAGGGT CATCAATGCT 1589
 AAGTATCTGT AACTATTTTA ACTCTCAAAA CTTGTGATAT ACAAAGTCTA AATTATTAGA 1649
 CGACCAATAC TTTAGGTTTA AAGGCATACA AATGAAACAT TCAAAAATCA AAATCTATTC 1709
 TGTTTCTCAA ATAGTGAATC TTATAAAATT AATCACAGAA GATGCAAATT GCATCAGAGT 1769
 CCCTTAAAT TCCTCTTCGT ATGAGTATTT GAGGGAGGAA TTGGTGATAG TTCCTACTTT 1829
 CTATTGGATG GTACTTTGAG ACTCAAAGC TAAGCTAAGT TGTGTGTGTG TCAGGGTGCG 1889
 GGGTGTGGAA TCCCATCAGA TAAAAGCAA TCCATGTAAT TCATTCAGTA AGTTGTATAT 1949
 GTAGAAAAAT GAAAAGTGGG CTATGCAGCT TGGAAACTAG AGAATTTTGA AAAATAATGG 2009
 AAATCACAAG GATCTTTCTT AAATAAGTAA GAAAATCTGT TTGTAGAATG AAGCAAGCAG 2069
 GCAGCCAGAA GACTCAGAAC AAAAGTACAC ATTTTACTCT GTGTACACTG GCAGCACAGT 2129
 GGGATTATTT TACCTCTCCC TCCCTAAAAA CCCACACAGC GGTTCCTCTT GGGAAATAAG 2189
 AGGTTTCCAG CCCAAAGAGA AGGAAAGACT ATGTGGTGTT ACTCTAAAAA GTATTTAATA 2249
 ACCGTTTGTG TGTGCTGTT GCTGTTTTGA AATCAGATTG TCTCCTCTCC ATATTTTATT 2309
 TACTTCATTC TGTTAATTC TGTGGAATTA CTTAGAGCAA GCATGGTGAA TTCTCAACTG 2369

TAAAGCCAAA TTTCTCCATC ATTATAATTT CACATTTTGC CTGGCAGGTT ATAATTTTTTA 2429
 TATTTCCACT GATAGTAATA AGGTAATAATC ATTACTTAGA TGGATAGATC TTTTTCATAA 2489
 AAAGTACCAT CAGTTATAGA GGAAGTCAT GTTCATGTTT AGGAAGGTCA TTAGATAAAG 2549
 CTTCTGAATA TATTATGAAA CATTAGTTCT GTCATTCTTA GATTCTTTTT GTTAAATAAC 2609
 TTTAAAAGCT AACTTACCTA AAAGAAATAT CTGACACATA TGAACCTTCTC ATTAGGATGC 2669
 AGGAGAAGAC CCAAGCCACA GATATGTATC TGAAGAATGA ACAAGATTCT TAGGCCCGGC 2729
 ACGGTGGCTC ACATCTGTAA TCTCAAGAGT TTGAGAGGTC AAGGCGGGCA GATCACCTGA 2789
 GGTCAGGAGT TCAAGACCAG CCTGGCCAAC ATGATGAAAC CCTGCCTCTA CTAATAATAC 2849
 AAAAATTAGC AGGGCATGGT GGTGCATGCC TGCAACCCTA GCTACTCAGG AGGCTGAGAC 2909
 AGGAGAATCT CTTGAACCCT CGAGGCGGAG GTTGTGGTGA GCTGAGATCC CTCTACTGCA 2969
 CTCCAGCCTG GGTGACAGAG ATGAGACTCC GTCCCTGCCG CCGCCCCCGC CTTCCCCCCC 3029
 AAAAAGATTC TTCTTCATGC AGAACATACG GCAGTCAACA AAGGGAGACC TGGTCCAGG 3089
 TGTCGAAGT ACTTATTTTC AGTAAATTAG CAATGAAAGA ATGCCATGGA ATCCCTGCC 3149
 AAATACCTCT GCTTATGATA TTGTAGAATT TGATATAGAG TTGTATCCCA TTTAAGGAGT 3209
 ACGATGTAGT AGGAAAGTAC TAAAAACAAA CACACAAACA GAAAACCCTC TTTGCTTTGT 3269
 AAGGTGGTTC CTAAGATAAT GTCAGTGCAA TGCTGGAAAT AATATTTAAT ATGTGAAGGT 3329
 TTTAGGCTGT GTTTTCCCCT CCTGTTCTTT TTTTCTGCCA GCCCTTTGTC ATTTTTCAG 3389
 GTCAATGAAT CATGTAGAAA GAGACAGGAG ATGAAACTAG AACCAGTCCA TTTTGCCCCT 3449
 TTTTTTATTT TCTGGTTTTG GTAAAAGATA CAATGAGGTA GGAGGTTGAG ATTTATAAAT 3509
 GAAGTTTAAT AAGTTTCTGT AGCTTTGATT TTTCTCTTTC ATATTTGTTA TCTTGATAA 3569
 GCCAGAATTG GCCTGTAAAA TCTACATATG GATATTGAAG TCTAAATCTG TTCAACTAGC 3629
 TTACACTAGA TGGAGATATT TTCATATTCA GATACACTGG AATGTATGAT CTAGCCATGC 3689
 GTAATATAGT CAAGTGTGTT AAGGTATTTA TTTTAAATAG CGTCTTTAGT TGTGGACTGG 3749
 TTCAAGTTTT TCTGCCAATG ATTTCTTCAA ATTTATCAA TATTTTTCCA TCATGAAGTA 3809
 AAATGCCCTT GCAGTCACCC TTCTGAAGT TTGAACGACT CTGCTGTTTT AAACAGTTTA 3869
 AGCAAATGGT ATATCATCTT CCGTTTACTA TGTAAGCTTA CTGCAGGCTT ACGCTTTTGA 3929
 GTCAGCGGCC AACTTTATTG CCACCTTCAA AAGTTTATTA TAATGTTGTA AATTTTACT 3989
 TCTCAAGGTT AGCATACTTA GGAGTTGCTT CACAATTAGG ATTCAGGAAA GAAAGAACTT 4049
 CAGTAGGAAC TGATTGGAAT TTAATGATGC AGCATTCAAT GGTACTAAT TTCAAAGAAT 4109
 GATATTACAG CAGACACACA GCAGTTATCT TGATTTTCTA GGAATAATTG TATGAAGAAT 4169
 ATGGCTGACA ACACGGCCTT ACTGCCACTC AGCGGAGGCT GACTAATGA ACACCTACC 4229
 CTTCTTTTCT TTTCTCTCAC ATTTCTATGAG CGTTTTGTAG GTAACGAGAA AATTGACTTG 4289
 CATTTGCATT ACAAGGAGGA GAAACTGGCA AAGGGGATGA TGGTGGAAGT TTTGTTCTGT 4349
 CTAATGAAGT GAAAAATGAA AATGCTAGAG TTTTGTGCAA CATAATAGTA GCAGTAAAAA 4409
 CCAAGTGAAA AGTCTTTCCA AAAGTGTGTT AAGAGGGCAT CTGCTGGGAA ACGATTTGAG 4469
 GAGAAGGTAC TAAATTGCTT GGTATTTTCC GTAG GA ACC CCA GAG CGA AAT ACA 4523

@@@@ Gly Thr Pro Glu Arg Asn Thr
 @@@@ 115

GTT TGC AAA AGA TGT CCA GAT GGG TTC TTC TCA AAT GAG ACG TCA TCT 4571
 Val Cys Lys Arg Cys Pro Asp Gly Phe Phe Ser Asn Glu Thr Ser Ser
 120 @@@@ 125 @@@@ 130 @@@@ 135

AAA GCA CCC TGT AGA AAA CAC ACA AAT TGC AGT GTC TTT GGT CTC CTG 4619
 Lys Ala Pro Cys Arg Lys His Thr Asn Cys Ser Val Phe Gly Leu Leu
 @@@@ 140 @@@@ 145 @@@@ 150

CTA ACT CAG AAA GGA AAT GCA ACA CAC GAC AAC ATA TGT TCC GGA AAC 4667
 Leu Thr Gln Lys Gly Asn Ala Thr His Asp Asn Ile Cys Ser Gly Asn
 @@@@ 155 @@@@ 160 @@@@ 165

AGT GAA TCA ACT CAA AAA TGT GGA ATA G GTAATTACAT TCCAAAATAC 4715
 Ser Glu Ser Thr Gln Lys Cys Gly Ile
 @@@@ 170 @@@@ 175

GTCTTTGTAC GATTTTGTAG TATCATCTCT CTCTCTGAGT TGAACACAAG GCCTCCAGCC 4775
 ACATTCTTGG TCAAACTTAC ATTTTCCCCT TCTTGAATCT TAACCAGCTA AGGCTACTCT 4835

09/03/67

CGATGCATTA CTGCTAAAGC TACCACTCAG AATCTCTCAA AAACATCATCT TCTCACAGAT 4895
AACACCTCAA AGCTTGATTT TCTCTCCTTT CACACTGAAA TCAAATCTTG CCCATAGGCA 4955
AAGGGCAGTG TCAAGTTTGC CACTGAGATG AAATTAGGAG AGTCCAAACT GTAGAATTCA 5015
CGTTGTGTGT TATTACTTTC ACGAATGTCT GTATTATTAA CTAAAGTATA TATTGGCAAC 5075
TAAGAAGCAA AGTGATATAA ACATGATGAC AAATTAGGCC AGGCATGGTG GCTTACTCCT 5135
ATAATCCCAA CATTTTGGGG GGCCAAGGTA GGCAGATCAC TTGAGGTCAG GATTTCAAGA 5195
CCAGCCTGAC CAACATGGTG AAACCTTGTC TCTACTAAAA ATACAAAAAT TAGCTGGGCA 5255
TGGTAGCAGG CACTTCTAGT ACCAGCTACT CAGGGCTGAG GCAGGAGAAT CGCTTGAACC 5315
CAGGAGATGG AGGTTGCAGT GAGCTGAGAT TGTACCACTG CACTCCAGTC TGGGCAACAG 5375
AGCAAGATTT CATCACACAC ACACACACAC ACACACACAC ACACATTAGA AATGTGTACT 5435
TGGCTTTGTT ACCTATGGTA TTAGTGCATC TATTGCATGG AACTTCCAAG CTACTCTGGT 5495
TGTGTTAAGC TCTTCATTGG GTACAGGTCA CTAGTATTAA GTTCAGGTTA TTCGGATGCA 5555
TTCCACGGTA GTGATGACAA TTCATCAGGC TAGTGTGTGT GTTCACCTTG TCACTCCCAC 5615
CACTAGACTA ATCTCAGACC TTCACTCAAA GACACATTAC ACTAAAGATG ATTTGCTTTT 5675
TTGTGTTTAA TCAAGCAATG GTATAAACCA GCTTGACTCT CCCCAAACAG TTTTTCGTAC 5735
TACAAAGAAG TTTATGAAGC AGAGAAATGT GAATTGATAT ATATATGAGA TTCTAACCCA 5795
GTTCCAGCAT TGTTTCATTG TGTAATTGAA ATCATAGACA AGCCATTTTA GCCTTTGCTT 5855
TCTTATCTAA AAAAAAAAAA AAAAAAATGA AGGAAGGGGT ATTAAAAGGA GTGATCAAAT 5915
TTTAACATTC TCTTTAATTA ATTCATTTTT AATTTTACTT TTTTTCATT ATTGTGCACT 5975
TACTATGTGG TACTGTGCTA TAGAGGCTTT AACATTATA AAAACACTGT GAAAGTTGCT 6035
TCAGATGAAT ATAGGTAGTA GAACGGCAGA ACTAGTATTC AAAGCCAGGT CTGATGAATC 6095
CAAAAACAAA CACCCATTAC TCCCATTTTC TGGGACATAC TTACTCTACC CAGATGCTCT 6155
GGGCTTTGTA ATGCCTATGT AAATAACATA GTTTTATGTT TGGTTATTTT CCTATGTAAT 6215
GTCTACTTAT ATATCTGTAT CTATCTCTTG CTTTGTTCCT AAAGGTAAAC TATGTGTCTA 6275
AATGTGGGCA AAAAATAACA CACTATTCCA AATTACTGTT CAAATTCCTT TAAGTCAGTG 6335
ATAATTATTT GTTTTGACAT TAATCATGAA GTTCCCTGTG GGTACTAGGT AAACCTTTAA 6395
TAGAATGTTA ATGTTTGAT TCATTATAAG AATTTTGGC TGTACTTAT TTACAACAAT 6455
ATTTCACTCT AATTAGACAT TTACTAACT TTCTCTGAA AACAATGCC AAAAAGAAG 6515
ATTAGAAGAC ACGTAAGCTC AGTTGGTCTC TGCCACTAAG ACCAGCCAAC AGAAGCTTGA 6575
TTTTATTCAA ACTTTGCATT TTAGCATATT TTACTCTGGA AAATTCAATT TGTTGTTT 6635
TTTGTTTTGT TTTGTATTGA ATAGACTCTC AGAAATCCAA TTGTTGAGTA AATCTTCTGG 6695
GTTTTCTAAG CTTTCTTTAG AT GTT ACC CTG TGT GAG GAG GCA TTC TTC AGG 6747

Asp Val Thr Leu Cys Glu Glu Ala Phe Phe Arg
180 185

TTT GCT GTT CCT ACA AAG TTT ACG CCT AAC TGG CTT AGT GTC TTG GTA 6795
Phe Ala Val Pro Thr Lys Phe Thr Pro Asn Trp Leu Ser Val Leu Val
190 195 200

GAC AAT TTG CCT GGC ACC AAA GTA AAC GCA GAG AGT GTA GAG AGG ATA 6843
Asp Asn Leu Pro Gly Thr Lys Val Asn Ala Glu Ser Val Glu Arg Ile
205 210 215

AAA CGG CAA CAC AGC TCA CAA GAA CAG ACT TTC CAG CTG CTG AAG TTA 6891
Lys Arg Gln His Ser Ser Gln Glu Gln Thr Phe Gln Leu Leu Lys Leu
220 225 230 235

TGG AAA CAT CAA AAC AAA GAC CAA GAT ATA GTC AAG AAG ATC ATC CAA G 6940
Trp Lys His Gln Asn Lys Asp Gln Asp Ile Val Lys Lys Ile Ile Gln
240 245 250

GTATGATAAT CTAAAAATAA AAGATCAATC AGAAATCAAA GACACCTATT TATCATAAAC 7000
CAGGAACAAG ACTGCATGTA TGTTTAGTTG TGTGGATCTT GTTTCCTGT TGGAATCATT 7060
GTTGGACTGA AAAAGTTTCC ACCTGATAAT GTAGATGTGA TTCCACAAAC AGTTATACAA 7120
GGTTTTGTTC TCACCCCTGC TCCCAGTTT CTTGTAAAG TATGTTGAAC ACTCTAAGAG 7180
AAGAGAAATG CATTGAAGG CAGGGCTGTA TCTCAGGGAG TCGTTCCAG ATCCCTTAAC 7240
GCTTCTGTAA GCAGCCCCTC TAGACCACCA AGGAGAAGCT CTATAACCAC TTTGTATCTT 7300

09/05/670

ACATTGCACC TCTACCAAGA AGCTCTGTTG TATTTACTTG GTAATTCTCT CCAGGTAGGC 7360
TTTTCGTAGC TTACAAATAT GTTCTTATTA ATCCTCATGA TATGGCCTGC ATTAAAATTA 7420
TTTTAATGGC ATATGTTATG AGAATTAATG AGATAAAATC TGAAAAGTGT TTGAGCCTCT 7480
TGTAGGAAAA AGCTAGTTAC AGCAAAATGT TCTCACATCT TATAAGTTTA TATAAAGATT 7540
CTCCTTTAGA AATGGTGTGA GAGAGAAACA GAGAGAGATA GGGAGAGAAG TGTGAAAGAA 7600
TCTGAAGAAA AGGAGTTTCA TCCAGTGTGG ACTGTAAGCT TTACGACACA TGATGGAAAG 7660
AGTTCTGACT TCAGTAAGCA TTGGGAGGAC ATGCTAGAAG AAAAAGGAAG AAGAGTTTCC 7720
ATAATGCAGA CAGGGTCAGT GAGAAATTCA TTCAGGTCCT CACCAGTAGT TAAATGACTG 7780
TATAGTCTTG CACTACCCTA AAAAATTCTA AGTATCTGAA ACCGGGGCAA CAGATTTTATG 7840
GAGACCAACG TCTTTGAGAG CTGATTGCTT TTGCTTATGC AAAGAGTAAA CTTTATATGTT 7900
TTGAGCAAAC CAAAAGTATT CTTTGAACGT ATAATTAGCC CTGAAGCCGA AAGAAAAGAG 7960
AAAATCAGAG ACCGTTAGAA TTGGAAGCAA CCAAATTCCC TATTTTATAA ATGAGGACAT 8020
TTTAACCCAG AAAGATGAAC CGATTTGGCT TAGGGCTCAC AGATACTAAG TGACTCATGT 8080
CATTAATAGA AATGTTAGTT CCTCCCTCTT AGGTTTGTAC CCTAGCTTAT TACTGAAATA 8140
TTCTCTAGGC TGTGTGTCTC CTTTAGTTCC TCGACCTCAT GTCTTTGAGT TTTAGATAT 8200
CCTCCTCATG GAGGTAGTCC TCTGGTGCTA TGTGTATTCT TTAAAGGCTA GTTACGGCAA 8260
TTAACTTATC AACTAGCGCC TACTAATGAA ACTTTGTATT ACAAAGTAGC TAACTTGAAT 8320
ACTTTCCTTT TTTTCTGAAA TGTTATGGTG GTAATTTCTC AAACCTTTTTC TTAGAAAAC 8380
GAGAGTGATG TGTCTTATTT TCTACTGTTA ATTTTCAAAA TTAGGAGCTT CTTCCAAAGT 8440
TTTGTGATGAT GCCAAAAATA TATAGCATAT TATCTTATTA TAACAAAAAA TATTTATCTC 8500
AGTTCTTAGA AATAAATGGT GTCACCTAAC TCCCTCTCAA AAGAAAAGGT TATCATTGAA 8560
ATATAATTAT GAAATCTGTC AAGAACCCTT TGCCTCACGC TTGTTTTATG ATGGCATTGG 8620
ATGAATATAA ATGATGTGAA CACTTATCTG GGCTTTTGTCT TTATGCAG AT ATT GAC 8676

Leu Cys Glu Asn Ser Val Gln Arg His Ile Gly His Ala Asn Leu Thr 255

CTC TGT GAA AAC AGC GTG CAG CGG CAC ATT GGA CAT GCT AAC CTC ACC 8724
Leu Cys Glu Asn Ser Val Gln Arg His Ile Gly His Ala Asn Leu Thr 255
255 260 265 270

TTC GAG CAG CTT CGT AGC TTG ATG GAA AGC TTA CCG GGA AAG AAA GTG 8772
Phe Glu Gln Leu Arg Ser Leu Met Glu Ser Leu Pro Gly Lys Lys Val
275 280 285

GGA GCA GAA GAC ATT GAA AAA ACA ATA AAG GCA TGC AAA CCC AGT GAC 8820
Gly Ala Glu Asp Ile Glu Lys Thr Ile Lys Ala Cys Lys Pro Ser Asp
290 295 300

CAG ATC CTG AAG CTG CTC AGT TTG TGG CGA ATA AAA AAT GGC GAC CAA 8868
Gln Ile Leu Lys Leu Leu Ser Leu Trp Arg Ile Lys Asn Gly Asp Gln
305 310 315

GAC ACC TTG AAG GGC CTA ATG CAC GCA CTA AAG CAC TCA AAG ACG TAC 8916
Asp Thr Leu Lys Gly Leu Met His Ala Leu Lys His Ser Lys Thr Tyr
320 325 330

CAC TTT CCC AAA ACT GTC ACT CAG AGT CTA AAG AAG ACC ATC AGG TTC 8964
His Phe Pro Lys Thr Val Thr Gln Ser Leu Lys Lys Thr Ile Arg Phe
335 340 345 350

CTT CAC AGC TTC ACA ATG TAC AAA TTG TAT CAG AAG TTA TTT TTA GAA 9012
Leu His Ser Phe Thr Met Tyr Lys Leu Tyr Gln Lys Leu Phe Leu Glu
355 360 365

ATG ATA GGT AAC CAG GTC CAA TCA GTA AAA ATA AGC TGC TTA 9054
Met Ile Gly Asn Gln Val Gln Ser Val Lys Ile Ser Cys Leu
370 375 380

4/10/82

TA	ACTGGAAA	TGGCCATTGA	GCTGTTTCCT	CACAATTGGC	GAGATCCCAT	GGATGAGTAA	9114
AC	TGTTTCTC	AGGCACTTGA	GGCTTTTCAGT	GATATCTTTC	TCATTACCAG	TGACTAATTT	9174
TG	CCACAGGG	TACTAAAAGA	AACTATGATG	TGGAGAAAGG	ACTAACATCT	CCTCCAATAA	9234
AC	CCCAAATG	GTTAATCCAA	CTGTCAGATC	TGGATCGTTA	TCTACTGACT	ATATTTTCCC	9294
TT	TATTACTGC	TTGCAGTAAT	TCAACTGGAA	ATTAAAAAAA	AAAAACTAGA	CTCCACTGGG	9354
CC	TTACTAAA	TATGGGAATG	TCTAACTTAA	ATAGCTTTGG	GATTCCAGCT	ATGCTAGAGG	9414
CT	TTTATTAG	AAAGCCATAT	TTTTTTCTGT	AAAAGTTACT	AATATATCTG	TAACACTATT	9474
AC	AGTATTGC	TATTTATATT	CATTCAGATA	TAAGATTTGG	ACATATTATC	ATCCTATAAA	9534
GA	AACGGTAT	GA	CTTAATTT	TAGAAAGAAA	ATTATATTCT	GTTTATTATG	9594
GAG	AAATATAT	ATATTTTTAA	TGGAAAGTTT	GTAGCATT	TTT	TCTAATAGGT	9654
TTT	TCTGTGT	GGAGTATTTT	TATAATTTTA	TCTGTATAAG	CTGTAATATC	ATTTTATAGA	9714
AA	ATGCATTA	TTAGTCAAT	TGTTTAATGT	TGGAAAACAT	ATGAAATATA	AATTATCTGA	9774
AT	ATTAGATG	CTCTGAGAAA	TTGAATGTAC	CTTATTTTAA	AGATTTTATG	GTTTTATAAC	9834
TAT	ATAAATG	ACATTATTA	AGTTTTCAAA	TTATTTT	TTA	TTGCTTTCTC	9894
ATT	TTT	TTT	TTT	TTT	TTT	TTT	9898

(2) INFORMATION FOR SEQ ID NO: 3:

- Sequence number: 3 → (i) SEQUENCE CHARACTERISTICS:
- (A) Length of sequence: 401
 - (B) Sequence type: amino acid
 - (C) Strandedness: single stranded
 - (D) Topology: linear
 - (E) Molecular type: protein

Sequence (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

Met	Asn	Asn	Leu	Leu	Cys	Cys	Ala	Leu	Val	Phe	Leu	Asp	Ile	Ser
-20														
Ile	Lys	Trp	Thr	Thr	Gln	Glu	Thr	Phe	Pro	Pro	Lys	Tyr	Leu	His
-5														
Tyr	Asp	Glu	Glu	Thr	Ser	His	Gln	Leu	Leu	Cys	Asp	Lys	Cys	Pro
10														
Pro	Gly	Thr	Tyr	Leu	Lys	Gln	His	Cys	Thr	Ala	Lys	Trp	Lys	Thr
25														
Val	Cys	Ala	Pro	Cys	Pro	Asp	His	Tyr	Tyr	Thr	Asp	Ser	Trp	His
40														
Thr	Ser	Asp	Glu	Cys	Leu	Tyr	Cys	Ser	Pro	Val	Cys	Lys	Glu	Leu
55														
Gln	Tyr	Val	Lys	Gln	Glu	Cys	Asn	Arg	Thr	His	Asn	Arg	Val	Cys
70														
Glu	Cys	Lys	Glu	Gly	Arg	Tyr	Leu	Glu	Ile	Glu	Phe	Cys	Leu	Lys
85														
His	Arg	Ser	Cys	Pro	Pro	Gly	Phe	Gly	Val	Val	Gln	Ala	Gly	Thr
100														
Pro	Glu	Arg	Asn	Thr	Val	Cys	Lys	Arg	Cys	Pro	Asp	Gly	Phe	Phe
115														
Ser	Asn	Glu	Thr	Ser	Ser	Lys	Ala	Pro	Cys	Arg	Lys	His	Thr	Asn
130														
Cys	Ser	Val	Phe	Gly	Leu	Leu	Leu	Thr	Gln	Lys	Gly	Asn	Ala	Thr
145														
His	Asp	Asn	Ile	Cys	Ser	Gly	Asn	Ser	Glu	Ser	Thr	Gln	Lys	Cys
160														
Gly	Ile	Asp	Val	Thr	Leu	Cys	Glu	Glu	Ala	Phe	Phe	Arg	Phe	Ala
175														
Val	Pro	Thr	Lys	Phe	Thr	Pro	Asn	Trp	Leu	Ser	Val	Leu	Val	Asp
190														
Asn	Leu	Pro	Gly	Thr	Lys	Val	Asn	Ala	Glu	Ser	Val	Glu	Arg	Ile
205														
Lys	Arg	Gln	His	Ser	Ser	Gln	Glu	Gln	Thr	Phe	Gln	Leu	Leu	Lys

please ensure all amino acids are from those shown in 1.822(b)(2) of sequence Rubric

220	Leu Trp Lys His Gln	225	Asn Lys Asp Gln Asp	230	Ile Val Lys Lys Ile
235	Gln Asp Ile Asp	240	Leu Cys Glu Asn Ser	245	Val Gln Arg His Ile
250	Gly His Ala Asn Leu	255	Thr Phe Glu Gln Leu	260	Arg Ser Leu Met Glu
265	Ser Leu Pro Gly Lys	270	Lys Val Gly Ala Glu	275	Asp Ile Glu Lys Thr
280	Ile Lys Ala Cys Lys	285	Pro Ser Asp Gln Ile	290	Leu Lys Leu Leu Ser
295	Leu Trp Arg Ile Lys	300	Asn Gly Asp Gln Asp	305	Thr Leu Lys Gly Leu
310	Met His Ala Leu Lys	315	His Ser Lys Thr Tyr	320	His Phe Pro Lys Thr
325	Val Thr Gln Ser Leu	330	Lys Lys Thr Ile Arg	335	Phe Leu His Ser Phe
340	Thr Met Tyr Lys Leu	345	Tyr Gln Lys Leu Phe	350	Leu Glu Met Ile Gly
355	Asn Gln Val Gln Ser	360	Val Lys Ile Ser Cys	365	Leu
370		375		380	

Sequence number: 4 (2) INFORMATION FOR SEQ ID NO: 4:

- (A) Length of sequence: 1206
 (B) Sequence type: nucleic acid
 (C) Strandedness: single stranded
 (D) Topology: linear
 (ii) Molecular type: cDNA

Sequence (X) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

ATGAACAAC	TGCTGTGCTG	CGCGCTCGTG	TTTCTGGACA	TCTCCATTAA	GTGGACCACC	60
CAGGAAACGT	TTCCTCCAAA	GTACCTTCAT	TATGACGAAG	AAACCTCTCA	TCAGCTGTTG	120
TGTGACAAAT	GTCCTCCTGG	TACCTACCTA	AAACAACACT	GTACAGCAAA	GTGGAAGACC	180
GTGTGCGCCC	CTTGCCCTGA	CCACTACTAC	ACAGACAGCT	GGCACACCAG	TGACGAGTGT	240
CTATACTGCA	GCCCCGTGTG	CAAGGAGCTG	CAGTACGTCA	AGCAGGAGTG	CAATCGCACC	300
CACAACCGCG	TGTGCGAATG	CAAGGAAGGG	CGCTACCTTG	AGATAGAGTT	CTGCTTGAAA	360
CATAGGAGCT	GCCCTCCTGG	ATTTGGAGTG	GTGCAAGCTG	GAACCCCAAG	GCGAAATACA	420
GTTTGCAAAA	GATGTCCAGA	TGGGTTCTTC	TCAAATGAGA	CGTCATCTAA	AGCACCCCTGT	480
AGAAAACACA	CAAATTGCAG	TGTCTTTGGT	CTCCTGCTAA	CTCAGAAAGG	AAATGCAACA	540
CACGACAACA	TATGTTCCGG	AAACAGTGAA	TCAACTCAAA	AATGTGGAAT	AGATGTTACC	600
CTGTGTGAGG	AGGCATTCTT	CAGGTTTGCT	GTTCTTACAA	AGTTTACGCC	TAAGTGGCTT	660
AGTGTCTTGG	TAGACAATTT	GCCTGGCACC	AAAGTAAACG	CAGAGAGTGT	AGAGAGGATA	720
AAACGGCAAC	ACAGCTCACA	AGAACAGACT	TTCCAGCTGC	TGAAGTTATG	GAAACATCAA	780
AACAAAGACC	AAGATATAGT	CAAGAAGATC	ATCCAAGATA	TTGACCTCTG	TGAAAACAGC	840
GTGCAGCGGC	ACATTGGACA	TGCTAACCTC	ACCTTCGAGC	AGCTTCGTAG	CTTGATGGAA	900
AGCTTACCGG	GAAAGAAAGT	GGGAGCAGAA	GACATTGAAA	AAACAATAAA	GGCATGCAAA	960
CCCAGTGACC	AGATCCTGAA	GCTGCTCAGT	TTGTGCGGAA	TAAAAAATGG	CGACCAAGAC	1020
ACCTTGAAGG	GCCTAATGCA	CGCACTAAAG	CACTCAAAGA	CGTACCACTT	TCCCAAAACT	1080
GTCACCTCAGA	GTCTAAAGAA	GACCATCAGG	TTCCTTCACA	GCTTCACAAT	GTACAAATTG	1140
TATCAGAACT	TATTTT TAGA	AATGATAGGT	AACCAGGTCC	AATCACTAAA	AATAAGCTGC	1200
TTATAA						1206

delete

Please consult sample sequence listing (attached) for valid format. Suggestion: consult sequence rules

(3) Computer: Apple Macintosh;
 (i) Operating System: Macintosh;
 (ii) Macintosh File Type: text with line termination
 (iii) Line Terminator: Pre-defined by text type file;
 (iv) Pagination: Pre-defined by text type file;
 (v) End-of-file: Pre-defined by text type file;
 (vi) Media: (A) Diskett—3.50 inch, 400 Kb storage;
 (B) Diskette—3.50 inch, 800 Kb storage;
 (C) Diskette—3.50 inch, 1.4 Mb storage;
 (vii) Print Command: Use PRINT command from any Macintosh Application that processes text files, such as MacWrite or Teach Text;
 (4) Magnetic tape: 0.5 inch, up to 2400 feet;
 (i) Density: 1600 or 6250 bits per inch, 9 track;
 (ii) Format: raw, unblocked;
 (iii) Line Terminator: ASCII Carriage Return plus optional ASCII Line Feed;
 (iv) Pagination: ASCII Form Feed or Series of Line Terminators;
 (v) Print Command (Unix shell version given here as sample response—mt/dev/rmt0; lpr/dev/rmt0);
 (g) Computer readable forms that are submitted to the Office will not be returned to the applicant.
 (h) All computer readable forms shall have a label permanently affixed thereto on which has been hand printed or typed, a description of the format of the computer readable form as well as the name of the applicant, the title of the invention, the date on which the data were recorded on the computer readable form and the name and type of computer and operating system which generated the files on the computer readable form. If all of this information cannot be printed on a label affixed to the computer readable form, by reason of size or otherwise, the label shall include the name of the applicant and the title of the invention and a reference number, and the additional information may be provided on a container for the computer readable form with the name of the applicant, the title of the invention, the reference number and the additional information affixed to the container. If the computer readable form is submitted after the date of filing

under 35 U.S.C. 111, after the date of entry in the national stage under 35 U.S.C. 371 or after the time of filing in the United States Receiving Office, an international application under the PCT, the labels mentioned herein must also include the date of the application and the application number, including series code and serial number.

§ 1.825 Amendments to or replacement of sequence listing and computer readable copy thereof.

(a) Any amendment to the paper copy of the "Sequence Listing" (§ 1.821(c)) must be made by the submission of substitute sheets. Amendments must be accompanied by a statement that indicates support for the amendment in the application, as filed, and a statement that the substitute sheets include no new matter. Such a statement must be a verified statement if made by a person not registered to practice before the Office.

(b) Any amendment to the paper copy of the "Sequence Listing," in accordance with paragraph (a) of this section, must be accompanied by a substitute copy of the computer readable form (§ 1.821(e)) including all previously submitted data with the amendment incorporated therein, accompanied by a statement that the copy in computer readable form is the same as the substitute copy of the "Sequence Listing." Such a statement must be a verified statement if made by a person not registered to practice before the Office.

(c) Any appropriate amendments to the "Sequence Listing" in a patent, e.g., by reason of reissue or certificate of correction, must comply with the requirements of paragraphs (a) and (b) of this section.

(d) If, upon receipt, the computer readable form is found to be damaged or unreadable, applicant must provide, within such time as set by the Commissioner, a substitute copy of the data in computer readable form accompanied by a statement that the substitute data is identical to that originally filed. Such a statement must be a verified statement if made by a person not registered to practice before the Office.

Appendix A—Sample Sequence Listing

(1) GENERAL INFORMATION:

(i) APPLICANT: Doe, Joan X. Doe, John Q.
 (ii) TITLE OF INVENTION: Isolation and Characterization of a Gene Encoding a Protease from *Paramecium* sp.
 (iii) NUMBER OF SEQUENCES: 2
 (iv) CORRESPONDENCE ADDRESS:
 (A) ADDRESSEE: Smith and Jones
 (B) STREET: 123 Main Street
 (C) CITY: Smalltown
 (D) STATE: Anystate
 (E) COUNTRY: USA
 (F) ZIP: 12345
 (v) COMPUTER READABLE FORM:
 (A) MEDIUM TYPE: Diskette, 3.50 inch, 800 Kb storage
 (B) COMPUTER: Apple Macintosh
 (C) OPERATING SYSTEM: McIntosh 5.0
 (D) SOFTWARE: MacWrite
 (vi) CURRENT APPLICATION DATA:
 (A) APPLICATION NUMBER: 09/999,999
 (B) FILING DATE: 28-FEB-1989
 (C) CLASSIFICATION: 999/99
 (vii) PRIOR APPLICATION DATA:
 (A) APPLICATION NUMBER: PCT/US88/99999
 (B) FILING DATE: 01-MAR-1988
 (viii) ATTORNEY/AGENT INFORMATION:
 (A) NAME: Smith, John A.
 (B) REGISTRATION NUMBER: 00001
 (C) REFERENCE/DOCKET NUMBER: 01-0001
 (ix) TELECOMMUNICATION INFORMATION:
 (A) TELEPHONE: (909) 999-0001
 (B) TELEFAX: (909) 999-0002
 (2) INFORMATION FOR SEQ ID NO: 1:
 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 954 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
 (ii) MOLECULE TYPE: genomic DNA
 (iii) HYPOTHETICAL: yes
 (iv) ANTI-SENSE: no
 (vi) ORIGINAL SOURCE:
 (A) ORGANISM: *Paramecium* sp.
 (C) INDIVIDUAL/ISOLATE: XYZZ
 (G) CELL TYPE: unicellular organism
 (vii) IMMEDIATE SOURCE:
 (A) LIBRARY: genomic
 (B) CLONE: Para-XYZ2/36
 (x) PUBLICATION INFORMATION:
 (A) AUTHORS: Doe, Joan X. Doe, John Q.
 (B) TITLE: Isolation and Characterization of a Gene Encoding a Protease from *Paramecium* sp.
 (C) JOURNAL: Fictional Series
 (D) VOLUME: I
 (E) ISSUE: 1
 (F) PAGES: 1-20
 (G) DATE: 02-MAR-1988
 (K) RELEVANT RESIDUES IN SEQ ID NO: 1: FROM 1 TO 954

BILLING CODE 3510-16-M

Please consult

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

ATCGGGATAG TACTGGTCAA GACCGGTGGA CACCGGTAA CCCCAGTTAA GTACCGGTTA	60
TAGGCCATTT CAGGCCAAAT GTGCCCAACT ACGCCAATTG TTTTGCCAAC GGCCAACGTT	120
ACGTTTCGTAC GCACGTATGT ACCTAGGTAC TTACGGACGT GACTACGGAC ACTTCCGTAC	180
GTACGTACGT TTACGTACCC ATCCCAACGT AACCACAGTG TGGTCGCAGT GTCCCAGTGT	240
ACACAGACTG CCAGACATTC TTCACAGACA CCCC ATG ACA CCA CCT GAA CGT CTC	295
Met Thr Pro Pro Glu Arg Leu	
-30	
TTC CTC CCA AGG GTG TGT GGC ACC ACC CTA CAC CTC CTC CTT CTG GGG	343
Phe Leu Pro Arg Val Cys Gly Thr Thr Leu His Leu Leu Leu Leu Gly	
-25 -20 -15	
CTG CTG CTG GTT CTG CTG CCT GGG GCC CAT GTGAGGCAGC AGGAGAATGG	393
Leu Leu Leu Val Leu Leu Pro Gly Ala His	
-10 -5	
GGTGGCTCAG CCAAACCTTG AGCCCTAGAG CCCCCCTCAA CTCTGTTCTC CTAG GGG	450
Gly	
CTC ATG CAT CTT GCC CAC AGC AAC CTC AAA CCT GCT GCT CAC CTC ATT	498
Leu Met His Leu Ala His Ser Asn Leu Lys Pro Ala Ala His Leu Ile	
1 5 10 15	
GTAAACATCC ACCTGACCTC CCAGACATGT CCCCACCAGC TCTCCTCCTA CCCCTGCCTC	558
AGGAACCCAA GCATCCACCC CTCTCCCCCA ACTTCCCCCA CGCTAAAAAA AACAGAGGGA	618
GCCCACTCCT ATGCCTCCCC CTGCCATCCC CCAGGAACTC AGTTGTTTCTG TGCCCACTTC	678
TAC CCC AGC AAG CAG AAC TCA CTG CTC TGG AGA GCA AAC ACG GAC CGT	726
Tyr Pro Ser Lys Gln Asn Ser Leu Leu Trp Arg Ala Asn Thr Asp Arg	
20 25 30	
GCC TTC CTC CAG GAT GGT TTC TCC TTG AGC AAC AAT TCT CTC CTG GTC	774
Ala Phe Leu Gln Asp Gly Phe Ser Leu Ser Asn Asn Ser Leu Leu Val	
35 40 45	
TAGAAAAAAT AATTGATTTC AAGACCTTCT CCCCATTCTG CCTCCATTCT GACCATTTC	834
GGGGTCGTCA CCACCTCTCC TTTGGCCATT CCAACAGCTC AAGTCTTCCC TGATCAAGTC	894
ACCGGAGCTT TCAAAGAAGG AATTCTAGGC ATCCCAGGGG ACCCACACCT CCCTGAACCA	954

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 82 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(ix) FEATURE:

(A) NAME/KEY: signal sequence

(B) LOCATION: -34 to -1

(C) IDENTIFICATION METHOD: similarity
to other signal sequences, hydrophobic(D) OTHER INFORMATION: expresses
protease

(x) PUBLICATION INFORMATION:

(A) AUTHORS: Doe, Joan X, Doe, John Q

(B) TITLE: Isolation and Characterization
of a Gene Encoding a Protease from
Paramecium sp.

(C) JOURNAL: Fictional Genes

(D) VOLUME: I

(E) ISSUE: 1

(F) PAGES: 1-20

(G) DATE: 02-MAR-1988

(K) RELEVANT RESIDUES IN SEQ ID NO:

2: FROM -34 TO 48

BILLING CODE 3510-10-M

Here's where sequence 2 starts (after
the sequence data of SEQ ID NO: 1:)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met Thr Pro Pro Glu Arg Leu Phe Leu Pro Arg Val Cys Gly Thr Thr
-30 -25 -20

Leu His Leu Leu Leu Leu Gly Leu Leu Leu Val Leu Leu Pro Gly Ala
-15 -10 -5

His Gly Leu Met His Leu Ala His Ser Asn Leu Lys Pro Ala Ala His
1 5 10

Leu Ile Tyr Pro Ser Lys Gln Asn Ser Leu Leu Trp Arg Ala Asn Thr
15 20 25 30

Asp Arg Ala Phe Leu Gln Asp Gly Phe Ser Leu Ser Asn Asn Ser Leu
35 40 45

Leu Val

BILLING CODE 3510-16-C